

Term Structure Forecasting: No-arbitrage Restrictions Versus Large Information set*

Carlo A. Favero^a Linlin Niu^b Luca Sala^{c†}

^{a,c}IGIER, Bocconi University

^bWISE, Xiamen University

February 2010

Abstract

This paper addresses the issue of forecasting the term structure. We provide a unified state-space modeling framework that encompasses different existing discrete-time yield curve models. Within such framework we analyze the impact of two modeling choices, namely the imposition of no-arbitrage restrictions and the size of the information set used to extract factors, on the forecasting performance. Using US yield curve data, we find that both no-arbitrage and large information help in forecasting but no model uniformly dominates the other. No-arbitrage models are more useful at shorter horizon for shorter maturities. Large information sets are more useful at longer horizons and longer maturities. We also find evidence for a significant feedback from yield curve models to macroeconomic variables that could be exploited for macroeconomic forecasting.

Keywords: Yield curve, term structure of interest rates, forecasting, large data set, factor models

JEL Classification: C33, C53, E43, E44

*We thank Andrew Ang, Michel van der Wel, Marco Taboga and seminar participants at Bocconi University, Norges Bank for useful comments, WISE Information Center for computational support.

^{†a,c}: IGIER, Università Bocconi, Via Guglielmo Röntgen, 20136 Milan, Italy. E-mail: carlo.favero@unibocconi.it, and luca.sala@unibocconi.it. ^b: WISE, Economics Building A301, Xiamen University, 361005, Xiamen, China. E-mail: llniu@xmu.edu.cn.

1 Introduction

Yields of maturities longer than one period are risk-adjusted averages of expected future short-rates. Short term rates are monetary policy instruments, controlled by central banks. Modeling and forecasting the term structure therefore requires modeling and forecasting risk as perceived by the market and modeling and forecasting future monetary policy rates. This paper provides a unified framework encompassing the many different approaches employed in the literature to model the term structure and provides empirical evidence on their forecasting performance for U.S. data.

The amount of information relevant to modeling and forecasting expected monetary policy and risk is potentially enormous, therefore the choice of a "parsimonious" specification capable of capturing all relevant information is the crucial step in the modeling strategy. Following this intuition, all the relevant information on pricing bonds at any given point in time is often summarized by a small number of factors. As a consequence, the task of forecasting the term structure is simplified to that of forecasting a small number of factors. Different modeling strategies are defined by the restrictions used to shrink the information available in a few factors (Litterman and Scheinkman, 1991). The traditional finance literature limits the information set to a number of observable yields and uses two alternative methods: extraction of latent factors via cross-sectional interpolation methods and extraction of latent factors by exploiting no-arbitrage restrictions.

Among the cross-sectional interpolation methods, the Nelson and Siegel (1987) approach is the most popular. The Nelson and Siegel three factor model explains most variances of yields at different maturities with a very good in-sample fit. Diebold and Li (2006) have successfully considered the out-of-sample forecasting performance of this model by assuming that the three factors follow AR(1) processes.

Among no-arbitrage models, the common approach is to assume a linear model for the latent factors and to restrict the factor loadings so as to rule out arbitrage strategies on bonds of different maturities. No-arbitrage restrictions serve not only for reducing the dimension of the parameter space, but also contribute to the theoretical consistency of the model. Dai and Singleton (2000) and Piazzesi (2003) have surveyed the specification issues of affine term structure models in continuous time and discrete time, respectively. Duffee (2002) has shown the usefulness of essentially affine term structure models ($A_0(3)$ ¹) in forecasting.

¹The symbol $A_0(n)$, used to denote essentially affine term structure models, refers to

These two approaches have been recently merged in an affine arbitrage-free Nelson-Siegel (AFNS) model, see Christensen, Diebold and Rudebusch (2007) or Le Grand (2007), where the traditional Nelson and Siegel structure is modified to rule out arbitrage opportunities.

Models mentioned above are traditionally based only on the information contained in the term structure. Financial markets are clearly not insulated from the rest of the economy. The feedback from the state of the economy to the short term interest rate is explicitly considered in the monetary policy reaction function introduced by Taylor (1993) and by now widely adopted to explain the behaviour of central banks.

Several papers indicate that macroeconomic variables have strong effects on future movements of the yield curve (among others, Ang and Piazzesi (2003), Diebold, Rudebusch, and Aruoba (2006) and Rudebusch and Wu (2008)). In particular, Ang and Piazzesi (2003) use an $A_0(3)$ model, and show that a mixed model (with three latent financial factors plus output and inflation) performs better than a yields-only model in terms of one step ahead forecast at monthly frequency.

One question that naturally arises in this context is how to efficiently summarize the large amount of macroeconomic information available. Factor models suited to deal with large cross-sections have therefore become increasingly popular in the forecasting literature. As shown in Stock and Watson (2002) and Forni, Hallin, Lippi, and Reichlin (2005), by decomposing large panels of time series in common and idiosyncratic components, information can be used efficiently, dimensionality greatly reduced and forecasting efficiency improved. Giannone, Reichlin and Sala (2004) show that a two dynamic factor model produces forecasting accuracy of the federal funds rate similar to that of the market.

We will set up an encompassing framework in which we will assess the relative importance of no-arbitrage restrictions versus large information sets in forecasting the yield curve.

We choose to evaluate alternative term structure models on the basis of their out-of-sample forecasting performance for different yields. In this way, we will have a uniform ground to compare models with very different features, settings and number of parameters.

We are not the first to provide evidence on the forecasting performance of alternative term structure models: recently Moench (2008) proposed a no-arbitrage factor-augmented VAR (FAVAR, see Bernanke, Boivin and Elias, 2005), in which financial factors are augmented with macroeconomic factors,

the fact that in the affine (" A ") model, there are " n " state variables, but none of the states drives the conditional variance of the state innovation, hence denoted by subscript " $_0$ ".

and compared its forecasting performance with a number of alternatives and showed that a no-arbitrage FAVAR delivers almost uniformly better forecasts at horizons from 6 to 12 months ahead. Our exercise differs from Moench's in many aspects. First, we propose a more general framework to evaluate systematically a larger set of models; second, we base our forecasting comparison exercise on a rolling window estimation with fixed size, in which parameters are re-estimated at each stage, while Moench considered a recursive estimation strategy and expanded the estimation window as new observations are included in the sample; third, none of the models we compare is our proposed model; fourth, our empirical results are different in that our evidence is not overwhelmingly in favor of the FAVAR model.

Concluding, we will also discuss the reverse issue of forecasting macroeconomic variables with term structure models.

The paper is organized as follows. In Section 2 we propose a unified state-space framework to evaluate the effects of incorporating factor information and/or no-arbitrage restrictions on the forecasting performance of empirical models of the yield curve. In Section 3 we describe the data. In Section 4 we discuss model specification and evaluate the forecasting performance of various models. Section 5 is devoted to the discussion of our empirical results. Section 6 discusses robustness issues and Section 7 concludes.

2 The general state-space representation

We study the dynamics of the term structure in the state-space model presented in equations (1) and (2). $y_{t,t+n}$ is the yield-to-maturity at time t of a zero-coupon bond maturing at time $t + n$. Yields with different maturities are collected in a vector $y_t = [y_{t,t+1}, y_{t,t+2}, \dots, y_{t,t+N}]'$. Equation (1) is the measurement equation, in which different yields $y_{t,t+n}$ are assumed to be determined by a set of state variables, collected in the vector X_t . Equation (2) is the state equation in which the states X_t are assumed to follow a VAR(1) process.

$$y_{t,t+n} = \frac{-1}{n} (A_n + B_n' X_t) + \varepsilon_{t,t+n} \quad \varepsilon_{t,t+n} \sim i.i.d.N(0, \sigma_n^2) \quad (1)$$

$$X_t = \mu + \Phi X_{t-1} + v_t \quad v_t \sim i.i.d.N(0, \Omega) \quad (2)$$

The variables in X_t can be either endogenous (that is, some of the elements of y_t could also be included in X_t) or exogenous, observable or latent.

The system composed of (1) and (2) is very general and can accommodate different specifications. Equation (1) illustrates how the yield curve is fitted.

This can be done by pure interpolation methods or by imposing no-arbitrage restrictions. When no-arbitrage restrictions are imposed, the entries of matrices A_n and B_n are constrained by cross-equation restrictions derived from economic theory. In equation (2) different specifications of the information set can be adopted according to the choice of what variables to include in X_t . As we shall see below, some models will include only factors extracted from yield curve data, while others will include a combination of factors from the yield curve and factors from macroeconomic data.

We take the forecasting performance as the metric to evaluate alternative models. We shall classify models along two dimensions: one will be based on the nature of the restrictions imposed on the measurement equation (1) - interpolation (or reduced form) versus no-arbitrage restrictions; the second will be based on the information set used to model the state dynamics - small versus large information set.

2.1 Interpolation versus no-arbitrage models

The number of models of the term structure available in the literature is vast. We select a limited number of models, each of which is to be considered as a representative of a class of models.

1. Diebold-Li model. In Diebold-Li (2006), three factors, extracted *à la* Nelson and Siegel (NS henceforth), are assumed to follow an unrestricted VAR. In this case we have:

$$B'_n = \left[-n, -\left(\frac{1 - e^{-\lambda n}}{\lambda}\right), -\left(\frac{1 - e^{-\lambda n}}{\lambda} - ne^{-\lambda n}\right) \right] \text{ and } A_n = 0$$

We denote the three yield factors as $NS_t = [NS_{1,t} \ NS_{2,t} \ NS_{3,t}]'$ and define $X_t = NS_t$. Equation (1) then takes the form:

$$y_{t,t+n} = NS_{1,t} + NS_{2,t} \left(\frac{1 - e^{-\lambda n}}{\lambda n}\right) + NS_{3,t} \left(\frac{1 - e^{-\lambda n}}{\lambda n} - e^{-\lambda n}\right) + \varepsilon_{t,t+n} \quad (3)$$

The dynamics of NS_t is assumed to follow an unrestricted VAR(1):

$$NS_t = \mu + \Phi NS_{t-1} + v_t \quad (4)$$

$NS_{1,t}$, $NS_{2,t}$, and $NS_{3,t}$ are estimated as parameters in a cross-section of yields, letting n , the maturity date, vary. In the time series dimension, $NS_{1,t}$, $NS_{2,t}$, and $NS_{3,t}$ have an immediate interpretation as latent factors. The loading on $NS_{1,t}$ is the only element in B'_n that does not decay to zero

as n tends to infinity; $NS_{1,t}$ can therefore be interpreted as the long-term factor, the level of the term structure. The loading on $NS_{2,t}$ is a monotone function that starts at 1 and decays to zero; $NS_{2,t}$ can be viewed as a short-term factor, the slope of the term structure. $NS_{3,t}$ is a medium term factor: its loading starts at zero, increases and then decays to zero, with the speed of decay determined by the parameter λ . This factor is usually interpreted as the curvature of the yield curve. Empirically, the first NS factor closely represents the 10 year yield; the second NS factor correlates well with the spread between long and short yields: $(10year - 1month)$; the third NS factor is close to: $(2 \times 2year - (10year + 3month))$, a measure of curvature of the yield curve.

This model will be considered as the benchmark in the class of unrestricted models.

2. No-arbitrage affine models, in which long yields are risk-adjusted expectations of average future short-rates and the coefficients of the state-space model are restricted so as to rule out arbitrage opportunities (see Appendix 1 for details). In this case, we follow the general discrete-time framework popularized after Ang and Piazzesi (2003). Defining the market price for risk associated with the state variables X_t as $\Lambda_t = \lambda_0 + \lambda_1 X_t$ and given the measurement equation of the short rate, $y_{t,t+1} = -(A_1 + B'_1 X_t) + \varepsilon_{t,t+1}$, it is possible to show that no-arbitrage imposes the following structure on the coefficients of the measurement equation (for $n \geq 1$):

$$\begin{aligned} A_{n+1} &= A_n + B'_n (\mu - \Omega \lambda_0) + \frac{1}{2} B'_n \Omega B_n + A_1 \\ B'_{n+1} &= B'_n (\Phi - \Omega \lambda_1) + B'_1 \end{aligned}$$

The restrictions imply that once the coefficients on the short rate equation (A_1, B'_1) are fixed, all the other coefficients for longer maturity yields are determined by the following equations:

$$\begin{aligned} B_{n+1} &= \left[\sum_{i=0}^n (\Phi' - \lambda'_1 \Omega)^i \right] B_1 \\ A_{n+1} &= (n+1)A_1 + \sum_{i=0}^n B^{(i)}, \text{ where } B^{(i)} = B'_i (\mu - \Omega \lambda_0) + \frac{1}{2} B'_i \Omega B_i. \end{aligned}$$

In this setup the state vector is assumed to be of dimension 3. Following Chen and Scott (1993), the states are extracted by inverting the measurement equation, assuming that exactly 3 yields are observed without error (see details of this method and the corresponding likelihood function in Appendix 2). The Chen-Scott factors are denoted by $CS_t = [CS_{1,t} \ CS_{2,t} \ CS_{3,t}]'$. We define $X_t = CS_t$.

3. Affine arbitrage-free Nelson-Siegel (AFNS) model as in Christensen, Diebold and Rudebusch (2007) and Le Grand (2007). This model imposes the Nelson-Siegel structure on the canonical representation of affine models, so that

$$B'_n = \left[-n, -\left(\frac{1 - e^{-\lambda n}}{\lambda}\right), -\left(\frac{1 - e^{-\lambda n}}{\lambda} - ne^{-\lambda n}\right) \right].$$

Under the risk-neutral measure, it can be shown that the autoregressive coefficient matrix of the VAR(1) for the states is:

$$\Phi^Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 - \lambda & \lambda \\ 0 & 0 & 1 - \lambda \end{pmatrix}.$$

In order to exclude arbitrage opportunities, the measurement equation for the yields has to be adjusted by a constant term. Hence, differently from the original Nelson-Siegel model, $A_n \neq 0$.

Although the no-arbitrage Nelson-Siegel (NANS) model is appealing as a unifying framework that links the traditional Nelson-Siegel model to the affine no-arbitrage term structure model, it is significantly restrictive, as it is only consistent with the presence of exactly three state variables and does not allow for the inclusion of combinations of state variables of different nature, such as observable macro variables. For this reason, we decided to consider this model only in Section 5.3, where we address robustness issues.

2.2 State information set: small versus large

A second taxonomy considers small versus large information sets. We will add to the factors extracted from the yields additional variables and we will compare their effectiveness in forecasting yields.

We define an information set as *small* if it contains only the yield factors and/or a small set of observed macroeconomic variables. The huge literature on Taylor rules has shown that a nominal and a real variable are both important in driving the dynamics of nominal interest rates. We select the annual CPI inflation rate (π_t) as the measure of inflation, and the annual growth rate of the Index of Industrial Production (IPg_t) as the measure of real activity.²

²We have also considered the unemployment rate as an alternative indicator of real activity. Unemployment is outperformed by the index of industrial production for forecasting purposes.

1. In the class of small information set models we consider the following specifications.

1a) Unrestricted case. In addition to the NS_t factors, we add $[\pi_t \quad IPg_t]$ as observable factors. The state dynamics is described by a five-variables VAR.

1b) Restricted case. In addition to the CS_t factors, we consider $[\pi_t \quad IPg_t]$ as observable factors; as in case 1a), the state dynamics is described by a five-variables VAR.

2. In the class of large information set models, we extract common factors from a large panel of macroeconomic variables ($N = 162$). We estimate factors by static principal components, as in Stock and Watson (2002) and we call them $mf_t = [mf_{1,t} \quad mf_{2,t} \dots mf_{k,t}]'$. We evaluate the forecasting performance of "large N " macroeconomic factors in the following specifications:

2a) Unrestricted case. The macro factors are added to the NS factors: $X_t = [NS'_t \quad mf'_t]'$.

2b) Restricted case. The macro factors are added to the CS factors: $X_t = [CS'_t \quad mf'_t]'$.

2c) The macro factors are used as explanatory variables in a "generalized" Taylor rule (see Bernanke and Boivin, 2003): $X_t = [y_{t,1} \quad mf'_t]'$, both in the unrestricted and restricted models.

2d) The states are the macro factors: $X_t = mf_t$

We will employ two to four macro factors in our analysis.

Before discussing the empirical results some remarks are in order. The state space representation is so general and flexible that it can accomodate very different specifications not mentioned within our limited context. The advantage of using such an encompassing framework will be immediate when reporting results along the two dimensions discussed.

There, we will clearly see what are the elements driving the results and what is the role of each of them. For example, by going from reduced form to no-arbitrage model, given the information set and a similar number of parameters, we will see the role of no-arbitrage restrictions on the the forecasting performance; by changing the risk price specification, given the information set, we will be able to study how sensitive the results are to the form of risk price; by adding different set of macro factors, while holding constant the number of yield factors, we will see how the macro factors affect the forecast performance.

3 Data and macroeconomic factors

Our basic data set consists of Bliss data set of zero-coupon equivalent US yields for the sample 1974:2-2003:9 at the following 11 maturities: 1-month, 3-month, 6-month, 9-month, 1-year, 2-year, 3-year, 4-year, 5-year, 7-year, and 10-year.

We extract macroeconomic factors from a panel of 162 US macro monthly time series for the sample 1974:2-2003:9. The data set is the same used in Giannone, Reichlin and Sala (2004). We have excluded nine interest rates from the original 171 series.

# of series	Categories	Transformation
21	IP indices	$\ln X_t - \ln X_{t-12}$
39	Labor market	$\ln X_t - \ln X_{t-12}$
17	Sales, consumption spending	$\ln X_t - \ln X_{t-12}$
12	Inventory and orders	$\ln X_t - \ln X_{t-12}$
22	Financial markets, money and loans	$\ln X_t - \ln X_{t-12}$
25	Price indices	$\ln X_t - \ln X_{t-12}$
3	Import & export	$\ln X_t - \ln X_{t-12}$
23	Capacity utilization and inventory indices, etc.	X_t

The common factors have been extracted from the macro panel as follows. First, the data are transformed to obtain stationarity. We take annual log-difference for the series that contain trends (production indices, price indices including asset prices, money stock, etc.) while series stationary by their nature (capacity utilization, sentiment indicators, etc.) are considered in levels.

Second, we estimate the factors by principal components (Stock and Watson (2002)). We rank the factors according to their explanatory power³ and consider up to the fourth in our analysis.

As reported in Table 1, the first four factors explain up to 68% of the total variance in the panel. In Figures 1 and 2 the two macro variables and the first four macro factors extracted are plotted. As can be seen, the dynamics of the first two factors are closely related to IP growth rate and CPI inflation rate respectively. The first factor highly correlates with output growth. The R-squares of a regression of various industrial production indices on the first

³This is different from Ng and Ludvigson (2006). They construct a composite factor by combining several common factors according to their in sample significance in explaining the bond risk premia. We have tried to rank the factors according to their contribution to R-squares of yields, but we did not find clear evidence suggesting that such a strategy improves out-of-sample forecasting of yields.

factors are higher than 0.9 as shown in Table 1. The second factor closely follows inflation: it explains around 80% of variations in annual growth rates of PPI crude materials, CPI housing and CPI services as shown in Table 1. The third and fourth factor are related to financial variables and to the effective exchange rate.

4 Specification, estimation and forecast evaluation

4.1 Specification and estimation

In all models considered, we assume that the state dynamics follow a VAR(1).

In the unrestricted models, we do not make restrictions on the parameters and use a two-step OLS to estimate the state and measurement equations.

In no-arbitrage restricted models with only latent yield factors, we assume that the factors have zero mean $\mu = 0$, and that the VAR coefficient matrix Φ is lower triangular, with $\Omega = I$. This is the most general identified representation for the class of essentially $A_0(3)$ models (Dai and Singleton (2000)). In addition, in the short rate equation, $y_{t,t+1} = -A_1 - B_1 X_t + \varepsilon_{t,t+1}$, we set $-A_1 = \bar{r} + B_1 \bar{X}$, where \bar{r} is the historical mean of the short rate. We use the Chen-Scott method (see Appendix 2 for details) and estimate the model with maximum likelihood.

In restricted models in which the state vector is assumed to be composed by CS_t yield factors and observable macro variables or factors, we use the specification proposed by Pericoli and Taboga (2008), where the VAR coefficient matrix Φ is left unrestricted and the following conditions need to be met:

1) the covariance matrix Ω is block diagonal with the block corresponding to the unobservable yield factor being identity, and the block corresponding to the observable factors being unrestricted, i.e.

$$\Omega = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \Omega^o \end{bmatrix};$$

2) the loadings on the factors in the short rate equation are positive, $-A_1 \geq 0$;

3) $X_0^u = 0$.

Early specifications of this model often impose zero restrictions on the VAR coefficient matrix Φ (in Ang and Piazzesi (2003), Φ is block diagonal). These assumptions impose strong restrictions on the interaction between yield factors and macro factors. We have tried various specifications in our forecasting exercise. It turns out is that the least restricted specification, despite its heavier parameterization, does not have inferior forecasting performance⁴. We hold to the general specification below. We use the Chen-Scott method and estimate the model with maximum likelihood.

In the restricted model with only observable states, we follow a two step procedure. We first estimate the VAR for the states, then, given $(\hat{\mu}, \hat{\Phi}, \hat{\Omega})$, we estimate the prices of risk, λ_0 and λ_1 .

For all no-arbitrage models, we estimate three specifications for risk prices.

- Constant prices of risk: $\lambda_0 \neq 0$ and $\lambda_1 = 0$.
- Time-varying prices of risk, $\lambda_0 \neq 0$, λ_1 diagonal. This assumption, employed in Ang, Bekaert and Wei (2007), together with a diagonal Ω implies that prices of risk are independent.
- Nonzero factor correlations through the matrix, Φ , and state-dependent market prices of risk, λ_1 , with λ_1 being a full $K \times K$ matrix, K denoting the number of state variables, as discussed in Dai and Singleton (2002).

In the restricted model with Nelson-Siegel factors, we estimate a correlated-factor AFNS model as in Christensen, Diebold and Rudebusch (2007). For parsimony we assume a diagonal variance-covariance matrix Ω , as in Le Grand (2007). In this case, the constant term in the measurement equation of yield takes the following form

$$A_n = \sigma_{11}^2 \frac{n^3}{6} + \sigma_{22}^2 \left[\frac{n}{2\lambda^2} - \frac{1 - e^{-\lambda n}}{\lambda^3} + \frac{1 - e^{-2\lambda n}}{4\lambda^3} \right] + \sigma_{33}^2 \left[\frac{n}{2\lambda^2} + \frac{ne^{-\lambda n}}{\lambda^2} - \frac{n^2 e^{-2\lambda n}}{4\lambda} - \frac{3ne^{-2\lambda n}}{4\lambda^2} - \frac{2(1 - e^{-\lambda n})}{\lambda^3} + \frac{5(1 - e^{-2\lambda n})}{8\lambda^3} \right]$$

⁴Additional empirical evidence is available upon request.

4.2 Forecast procedure

We obtain h -steps ahead forecasts for the states by iterating the one-step model forward⁵:

$$\hat{X}_{t+h|t} = \sum_{i=0}^h \hat{\Phi}^i \hat{\mu} + \hat{\Phi}^h X_t \quad (5)$$

Forecasts based on different specifications are computed as follows:

4.2.1 Unrestricted models

1. Diebold and Li (2006). We obtain the Nelson and Siegel factors from equation (3). We fix λ , the parameter governing the speed of decay in the exponential function, at 0.0609, as calibrated in Diebold and Li (2006)⁶.

After having extracted the factors and estimated the unrestricted VAR(1), we obtain forecasts by iterated projections:

$$\hat{N}S_{t+h|t} = \sum_{i=0}^h \hat{\Phi}^i \hat{\mu} + \hat{\Phi}^h \hat{N}S_t \quad (6)$$

by using the NS parameterization:

$$\hat{y}_{t+h|t} = \hat{N}S_{1,t+h|t} + \hat{N}S_{2,t+h|t} \left(\frac{1 - e^{-\lambda n}}{\lambda n} \right) + \hat{N}S_{3,t+h|t} \left(\frac{1 - e^{-\lambda n}}{\lambda n} - e^{-\lambda n} \right)$$

2. Diebold-Li plus macro variables/factors. The Nelson-Siegel factors are extracted from yields as before. The state vector becomes: $X_t = [NS_t \ z_t]'$, where z_t contains the macro information and is modeled as a VAR(1). In this case, following Diebold, Rudebusch and Arouba (2006), we assume that the factor loadings of the yields on z_t in the measurement equations are zero. This specification is in line with the view that only three factors are needed to model the yield curve.

3. Interest rate rule-type VAR in which the state equation is unrestricted. In this setting, the yields are directly projected onto the states. Both the measurement and the state equations are estimated with OLS.

⁵The alternative would be to obtain forecasts by projecting h -step ahead: $\hat{X}_{t+h|t} = \hat{\mu}_h + \hat{A}_h X_t$. Given the nature of no-arbitrage models, only iterated forecast can be computed for them. For this reason, we employ iterated forecasts for all models. In Section 5.3 we check the robustness of our results to this choice.

⁶The factors extracted are insensitive to the choice of λ . A robustness check, in which λ is estimated, is presented in Section 5.3.

Unrestricted models are estimated with a two step OLS estimator, as in Diebold and Li (2006). Since these models are written in state space form, a one-step maximum likelihood (ML) estimation can be derived using the Kalman filter, as in Diebold, Rudebusch and Arouba (2006). As a robustness check, we estimate unrestricted Nelson and Siegel models with MLE and provide the result in Section 5.3.

4.2.2 No-arbitrage models

Forecasts in no-arbitrage models are obtained by using equation (5) in which the parameters are subject to the no-arbitrage restrictions.

4.3 Forecast comparison

We set the sample size fixed at 180 periods. We use rolling estimation by moving the sample forward by one observation at a time and re-estimating the model at every step, starting from the sample period 1978:1-1992:12. We consider four forecasting horizons (denoted by h): 1 month, 6 months, 12 months, and 24 months. For the 1-month ahead forecasting horizon, we conduct our exercise for all dates in the period 1993:1 - 2003:9, a total of 129 periods; for the 6 month ahead forecast, we end up with a total of 124 forecasts, and so on, up to the 24 month ahead forecast, for which we end up with 106 forecasts.

We choose two measures of forecasting performance. One is the ratio of the forecast root mean squared error (FRMSE) of each model to the FRMSE of a random walk forecast. We show the comparison of forecasting results from different models in Table 2A. The table shows better forecasts with respect to the random walk with bold characters for the range of $[0.9, 1)$, with added shading background for the range of $[0.8, 0.9)$, and with added underline for ratios smaller than 0.8.

FRMSE ratio tells the relative accuracy of each model for each maturity-horizon forecast compared to the random walk. Forecast errors originate from two sources: errors in forecasting yield levels, and errors in forecasting changes in yields. The FRMSE does not distinguish between these two types of errors. When models are subject to structural change and/or instability in the mean, even if they can forecast changes in yields relatively well, the forecasted levels might deviate from the realised data persistently so that the FRMSE is large. To take care of this problem, we complement the information in the FRMSE with a Bayesian probability indicator which

rewards models that better predict yield changes and controls for biases in the forecasts.

We construct this indicator by taking observed yields as the dependent variable in a regression. We use as regressors forecasts from two competing models at a time: one from one of the \bar{k} models discussed above and one from the random walk model. We assess the posterior probability of being included in the regression for each of the two forecasts. We then calculate the ratio of the posterior probabilities of the selected model forecast to be included in the general regression to that of the random walk forecast. We repeat the above procedure for all the models discussed above. Let us go into the details of the procedure (a more general discussion can be found in Koop, 2003).

Define $y = (y_1, \dots, y_t, \dots, y_T)'$ the realised yield at a specific maturity, where T is the total number of forecast.

We similarly define $\hat{y}_k = (\hat{y}_{k,1}, \dots, \hat{y}_{k,t}, \dots, \hat{y}_{k,T})'$ as the vector collecting the forecasts from model k .

We consider two forecasts at a time, the random walk forecast, \hat{y}_{RW} , and the forecast from model k , \hat{y}_k and define $X_k = [\hat{y}_{RW} \quad \hat{y}_k]$ as a $T \times 2$ matrix which contains the two forecasts of y at a specific forecast horizon h^7 . Given X_k , we can define three nested models M_r ($r = 1, 2, 3$) where M_3 is the most general one⁸:

$$M_1: \quad y = \alpha + \hat{y}_{RW}\beta_1 + \varepsilon \quad (7)$$

$$M_2: \quad y = \alpha + \hat{y}_k\beta_2 + \varepsilon \quad (8)$$

$$M_3: \quad y = \alpha + \hat{y}_{RW}\beta_1 + \hat{y}_k\beta_2 + \varepsilon \quad (9)$$

α is the intercept and ε is a $T \times 1$ vector of errors, which is assumed to be a priori distributed as $N(0_T, h^{-1}I_T)$. We assume all nested models have the same prior probability $p(M_r) = 1/3$.

This also implies that the prior for each regressor being included in the general model is equal. For each model with \tilde{k} regressors, we impose a relatively non-informative Normal-Gamma conjugate prior on the coefficients:

$$\begin{aligned} \gamma &\sim N(\underline{\gamma}, h^{-1}\underline{V}) \\ h &\sim \Gamma(\underline{s}^{-2}, \underline{v}) \end{aligned}$$

⁷We do not explicitly write the index h to save on notation.

⁸There are in principle 2^2 possible subsets of X_k . We exclude the empty set and focus on those subsets which contains at least one yield model forecast. For our comparison, whether or not to include the empty set will not affect the results. Excluding it speeds up the computations.

where $\gamma = (\alpha, \beta)$, by assigning:

$$\underline{\alpha} = 0, \underline{\beta}_i = 1/\tilde{k}$$

$$\underline{V} \text{ diagonal, } \underline{V}(1, 1) = 25 \text{ and } \underline{V}(j, j) = 1 \ (j > 1);$$

and

$$\underline{s}^{-2} = 0.5 \text{ and } \underline{v} = 4$$

These priors are chosen as such so that the implied prior variance of α is 100, and the implied prior variance of β_i is 4. The prior means of α and β_i imply that in each nested model the coefficients of different forecasts sum up to one and there is no persistent bias.

For each of the three nested models we calculate the posterior probability:

$$p(M_r|y) = \frac{p(y|M_r)p(M_r)}{\sum_{r=1}^3 p(y|M_r)p(M_r)}.$$

By integrating out the model posteriors, we can obtain the posterior probability that regressor \hat{y}_k is contained in the forecast regression model:

$$p(I_k = 1) = \sum_{r=1}^3 I_{k,r} p(M_r|y).$$

Our measure, the Bayesian Model Averaging Indicator (BMAI), will be the ratio:

$$BMAI_{k,RW} = \frac{p(I_k = 1)}{p(I_{RW} = 1)}$$

which can be written as:

$$BMAI_{k,RW} = \frac{p(M_2|y) + p(M_3|y)}{p(M_1|y) + p(M_3|y)}$$

If the ratio is bigger than 1, it means that regressor \hat{y}_k has a higher probability than \hat{y}_{RW} to be included in the forecasting regression. In Table 2B we report results. We highlight the ratio with bold number when it is bigger than 2, and add shaded background color when it is bigger than 5⁹, i.e. the selected model forecast is twice or five times more likely to be included in the regression than random walk. We make the number italic if the ratio is less than 0.5, i.e. it is twice likely that the random walk forecast will be included in the regression than the selected model.

⁹We have checked the robustness of these results to a range of different parameterizations for the relevant prior distributions.

When α is set to 0 and β_i is assumed to sum to 1, the priors of the model in equation (9) is similar to a standard forecast comparison regression, as in Stock and Watson (1999):

$$y = \lambda \hat{y}_1 + (1 - \lambda) \hat{y}_2 + \varepsilon. \quad (10)$$

The regression in equation (10) is too restrictive if the forecasts are biased, which may be the case for some yield models. In such occasion, the indicator penalizes the forecast with higher systematic bias, attributing little weight to it, even if the model can forecast well the changes of the variables. By allowing α to be different from zero, we take into account the biases in the forecasts and identify models with persistent bias but produce high correlation between forecasts and actual yields.

5 Empirical Results

We present our empirical results by discussing first our main evidence on yield curve forecasting. We identify the trade-offs of no-arbitrage restrictions between forecasting short versus longer maturity yields, and of macro factors between forecasting short versus longer horizons. We then use our framework to investigate whether yield curve models are useful to forecasting macroeconomic variables.

5.1 Yield curve forecast: no-arbitrage and/or large information set?

We report our main results on the forecasting performance in Table 2A for the FRMSE ratio and in Table 2B for the BMAI, respectively. In each table, we select 16 representative models and put them into 6 rows and 3 columns according to their characteristics. We report each model's forecasting performance in one sub-table along yield maturities (3, 12, 36, 60 and 120 months) and forecast horizons (1, 6, 12 and 24 months). Above each sub-table, we indicate the state vector.

The models are compiled along two dimensions.

Horizontally, we compare reduced form versus no-arbitrage models. Among no-arbitrage models we present the results of two specifications: constant risk price and time-varying risk price with λ_1 being a diagonal matrix. We do not list the results of time-varying risk price with full λ_1 , because the results of this specification are the worst performing among the three¹⁰. Vertically,

¹⁰Interested readers can refer to our appendix file, available upon request, for related results.

we compare small information set models with large information set ones. For the class of small information models, we present the three-yield factors model and the three-yield factor model augmented with macro variables. For the class of large information set, we present a model with three yield factors and macro factors, a model with one yield factor (the short rate) and macro factors and a model with macro factors only. We can therefore compare how macro factors fare compared with macro variables, and what is the relative role of yield curve versus macroeconomic information.

Among these models, some are existing in the literature, others are similar to existing models with minor differences in their specifications. If we denote the model in row m and column n by (m, n) , then model (1,1) is the Diebold-Li model with VAR(1) states; models (1,2) and (1,3) are $A_0(3)$ models with constant risk price and time-varying risk price respectively; model (2,1) is similar to the yield-macro model of Diebold, Rudebusch and Aruoba (2006) but with a different set of macro variables; models (2,2) and (2,3) are similar to Ang and Piazzesi (2003), with inflation and IP growth as explicit macro variables, allowing for the interaction between the dynamics of yield and macro variables. Models in the third and fourth rows have three yield factors plus macro factors extracted from large macro panel. Models relative to entries (4,2) and (4,3) are not reported; they have many parameters and their forecasting performance is not satisfactory. Models (5,2) and (5,3) are Moench (2008) type models; model (5,1) is a reduced form version of it. We put in the sixth and last row models with macro factors only, to complement the information in the first row, in which models with only yield factors are presented. This gives us an interesting comparison on the information content in yield curve prediction.

Let us now proceed to examine our results through several routes. We will start from the Diebold-Li model and move along the two dimensions aforementioned.

5.1.1 Reduced form versus no-arbitrage restrictions.

Moving from left to right in the first row of Table 2A, we find that in forecasting the short rate all three models have similar performance and they all beat the random walk model at all horizons. Compared with the Diebold-Li model, the no-arbitrage models with constant risk price has better forecast for medium and long term yields at all forecast horizons.

The forecasting performance of the no-arbitrage model with time-varying risk prices deteriorates, the more so, the longer the maturity: no-arbitrage restrictions help to reduce the FRMSE as long as the risk price is constant.

When we look at the second row of Table 2A where macro variables are

added to the three yield factors, the same pattern remains.

Does this findings dismiss the usefulness of time-varying risk price in forecasting yields? Not necessarily. We will go back to this issue when discussing the forecasting performance as measured by the BMAI criterion below.

5.1.2 Small versus large information set.

Moving downwards the first column of Table 2A, we observe that the model in entry (2,1), in which the inflation rate and the IP growth rate are added to the three Nelson-Siegel factors, does not improve upon Diebold-Li in terms of FRMSE.

When two latent macro factors are introduced (as in model (3,1)), there is not much improvement with respect to Diebold-Li. If compared with model (2,1), however, there is a clear tendency of improvement toward yields with longer maturities at longer forecast horizons. When three Nelson-Siegel factors are augmented with four macro factors as in model (4,1), in spite of the larger number of parameters, forecasts at 6 to 24 month horizons of all yields improve when compared with model (2,1), and forecast for short and medium term yields at 12 to 24 month horizons are better than both the Diebold-Li model and the random walk. This indicates that factors extracted from the large macro panel seem to better capture the real and nominal dimensions of the economy than the observable macro variables (as in Giannone, Reichlin and Sala, 2004).

Moving to model (5,1) and (6,1) where macro factors are predominant with respect to yield curve factors, the advantage in forecasting short and medium term yields at 12 to 24 month horizons remains, while the FRMSE ratios increase substantially at short horizons across the yield curve as the number of yield factors decreases. Kim (2007) discusses a similar finding when yield factors are replaced with observable macroeconomic variables.

We detect a clear pattern: macro information tends to improve yield forecasts at longer horizons; the macro factors extracted from the large macro panel have robust forecasting power at 12 to 24 months ahead, and this is true even in a VAR without yield curve factors. While previous research such as Ang and Piazzesi (2003) and Diebold, Rudebusch and Aruoba (2006) find evidence for the importance of macroeconomic information in forecasting the yield curve, our results are even clearer, thanks to the design of our experiments, in which the modification of one element at a time allows to identify the specific role of additional information .

5.1.3 No-arbitrage restrictions versus large information set.

Moving from the Diebold-Li model to either direction on Table 2A, both the no-arbitrage restrictions and large information set have some value added in forecasting the yield curve. Can we explore both at the same time? The block of "no-arbitrage restricted" and "large N " models in Tables 2A and 2B gives us some hints.

Model (3,2) is a no-arbitrage restricted model with constant risk price, with 3 latent yield factors and 2 macro factors. The restrictions reduce the FRMSE ratios for medium and long term yields compared to model (3,1) which has the same information set. For medium and long term yields at 12 and 24 month ahead forecast horizons, the large information set improves upon the no-arbitrage models (1,2) and (2,2) with small information set. This model enjoys the benefits of both imposing no-arbitrage restrictions and expanding the information set.

As we introduce time-varying risk price in model (3,3), although the short rate forecast in term of FRMSE ratios tends to improve for all horizons, the forecast of medium and long term yields deteriorates.

Moench (2008) takes a different approach to combine no-arbitrage with macro factors; he substitutes latent factors extracted from the yield curve with latent macro factors and keeps the short rate as the single yield factor. The VAR for the states can then be interpreted as a "generalised Taylor rule". Models (5,2) and (5,3) are simplified versions of Moench models, with less parameters in the risk price equations and less lags in the VAR. The pattern of the forecast performance of these models is similar, i.e. the long end of the yield curve is poorly predicted, but at longer forecast horizons the FRMSE ratio is low.

Comparing the three models in the fifth row, the no-arbitrage models fare worse than the unrestricted one; among the no-arbitrage models, time-varying risk price tends to improve the FRMSE compared to the constant risk price model. While the model in Moench (2008) did a better job at the long end of the yield curve, with a specification of more lags in the VAR, different time-varying risk specification and different period chosen, our experiment shows that this type of no-arbitrage models does not improve upon their corresponding reduced form model.

Models (6,2) and (6,3) with pure macro factors under no-arbitrage restrictions strengthen the above findings with further evidence.

Let us now turn to discuss the evidence on the BMAI indices in Table 2B. Models are collected in the same order as of Table 2A.

In the first row of Table 2B, the BMAI ratios are mostly around 1 for all three models compared with random walk. However, among the two no-arbitrage models, the time-varying risk price specification does better than the constant risk price model at 2 year horizon for short to medium maturity yields but not for the long end of the curve where the BMAI ratio falls below 0.5. In the second row of Table 2B, the constant risk price no-arbitrage model does not outperform neither the random walk nor the reduced form model, but the time-varying risk price model does better than the random walk at 1 to 2 year horizons again for the short to medium maturity yields, and better than the reduced form model at these horizons. The evidence of a better performance of time-varying risk price models is still present for the short to medium term yields when we move down to the third row when macro factors are added to the state vector. But the inferior forecast at the long maturity of the 10 year yield is also clear.

Moving downward the first column of Table 2B, when the three Nelson-Siegel factors are augmented with the inflation rate and IP growth rate, there is improvement in the BMAI ratios in the 24 month ahead forecasts for short to medium term yields, with respect to the Diebold-Li model. When two macro factors are added to the model (3,1), there is general improvement in the 12 month ahead forecasts. When four macro factors are augmented with three Nelson-Siegel factors, the forecasts at 12 to 24 month horizons of short and medium yields improve when compared to the model (2,1), although this improvement does not extend to long term yields. Moving to models (5,1) and (6,1) where four macro factors dominate the state vector, the improvement in forecasting short and medium term yields at 12 to 24 month horizons remains. The advantage of macro factors in these forecast horizons are also evident from the corresponding no-arbitrage models in the second and third column of these two rows. But the noisy forecast that had a high FRMSE ratios are not punished by the BMAI ratio at short horizons.

Overall, our results can be summarized as follows:

- (a) no-arbitrage restrictions with constant risk price in a three yield factor model generate low FRMSE;
- (b) macroeconomic factors extracted from large dataset are useful at predicting the yield curve at longer horizons of one to two years;
- (c) among no-arbitrage models, the FRMSE is systematically high for the longer maturities. Time-varying risk price is useful in capturing movements of the yield curve at short to medium maturities;
- (d) in models in which the state vector is composed of observable macro factors with no or little information on the yield curve, yields of longer ma-

turities are poorly predicted; the forecasts worsen when no-arbitrage restrictions are imposed.

5.2 Forecasting macroeconomic variables with yield curve information

Our framework not only allows to investigate the issue of yield curve forecast, but provides a laboratory to study whether models of the yield curve are useful to forecasting macroeconomic variables. Within this framework, we can examine: 1) whether the information on the term structure can contribute to the forecast of macro variables compared to a simple AR(1) time series model; 2) how sensitive are the results to different specifications for risk prices in no-arbitrage models; 3) the relative performance of no-arbitrage models and reduced form models; 4) the relative performance of Nelson-Siegel yield factors models and models with one observable yield factor (1-month rate).

Tables 3A and 3B report FRMSE ratios and BMAIs for macro variables. Column one contains results from the unrestricted models with three Nelson-Siegel yield factors. Column two shows forecasts from unrestricted models with the short rate as the single yield factor. Column three contains results from no-arbitrage models with 3 latent yield factors under three specifications of risk prices, i.e. "trp = 0" denotes constant risk price, "trp = 1" time-varying risk price with λ_1 being a diagonal matrix and "trp = 2" time-varying risk price with λ_1 a full matrix.

1) The FRMSE ratios in Table 3A show that when compared to a forecast from a simple time series AR(1), term structure information does not contribute to the predictability of inflation, but does increase the forecastability of real activity such as the IP growth. This is largely consistent with Ang, Bekaert and Wei (2007) for inflation forecast, and the findings of Ang, Piazzesi and Wei (2006) for GDP forecast. The BMAIs in Table 3B show that the inflation forecast 24 month ahead from no-arbitrage models might contain more useful information than the AR(1) model.

2) Within the no-arbitrage models, the results are sensitive to the risk price settings. We compare three risk price specifications in the third column. A parsimonious modeling is more favorable in terms of lower FRMSE ratios and higher BMAIs, especially for inflation, and the more so when forecast horizon increases. Fully time-varying risk prices are in general worse than the other two cases, except for the FRMSE of IP growth as in the model of entry (3,3) in Table 3A where IP growth is augmented with three yield factors. Although time-varying risk price is likely to capture patterns of time-varying risk premia of yields, it is more important for yield dynamics than

for macro variables. On the other hand, parameter uncertainty increases when λ_1 is time-varying and the forecast of macroeconomic variables quickly deteriorates.

3) By comparing the first and third columns we can compare the performance of reduced form Nelson-Siegel factor models to that of no-arbitrage latent factors models. The results show that the no-arbitrage models have better forecasting performance for inflation, especially for medium to long horizons, as long as the risk price setting is parsimonious. For IP growth, the reduced form models do a better job in FRMSE ratios while the BMAs are not conclusive.

4) Models in the second column has one yield factor - the short rate (1-month yield). Compared with the first column, results indicate that the information contained in the slope and curvature factors contributes to the forecast of IP growth - the measure of real activity. The evidence from BMAI ratios is mixed.

6 Robustness checks on model specification and forecasting procedure

In this Section, we discuss several issues related to the specification and use of Nelson-Siegel models.

6.1 Two-step OLS versus one-step MLE for unrestricted NS models

In previous Sections, we have estimated reduced form NS models with two-step OLS, calibrating the parameter λ . Here we check the robustness of results when the models are estimated with ML and when λ is estimated.

In the first row, first column of Table 4A, we show FMRSE for the Diebold-Li model estimated with ML with λ fixed at 0.0609. In the second row, first column of Table 4A, we show results when the model and λ are estimated with ML. The results are fairly similar to those in Table 2A; while the two-step OLS produces slightly lower FRMSE ratio with respect to the random walk for a majority of the yield-horizon combinations, the one-step MLE delivers somewhat better results towards the long end of yields for long forecast horizons.

Therefore, the two-step OLS procedure with fixed λ is robust to the estimation procedure and the choice of λ .

In the second column, we report results for the arbitrage free Nelson-Siegel

(AFNS) model proposed in Christensen, Diebold and Rudebusch (2007). Our implementation follows Le Grand, 2007 and assumes a diagonal variance-covariance matrix Ω . In the first row we calibrate λ , in the second row, we estimate it. Compared to the counterpart reduced form models in the first column, the no-arbitrage Nelson-Siegel has some advantage for the long yields at 24 month horizon, but in the rest of the combinations, the differences are minor.

Table 4B displays forecast results for the Nelson-Siegel model augmented with macro variables/factors, estimated with one-step ML. Results are similar to the corresponding two-step OLS forecast reported in Table 2A. Although in the Nelson-Siegel factor model with inflation and IP growth, the FRMSE ratios are slightly reduced for 12 to 24 month forecast horizons, in the rest of the cases ML produces slightly worse results than the two-step OLS.

6.2 Nelson-Siegel model performance and forecast periods

The results in Diebold and Li (2006) favored a specification in which each Nelson-Siegel factor was modelled as an AR(1) process. In our exercises, the VAR(1) specification does generally better than the AR(1). This subsection investigates the causes of this discrepancy.

Our exercise differs from that of Diebold and Li (2006) in several aspects: 1) we use iterated forecast for all models, while they use dynamic forecast, i.e. a regression of factors at $t + h$ on factors at t ; 2) we use a fixed-length rolling sample of 180 points, while they do a recursive forecast adding one observation at a time from the sample 1985:1-1994:1, through the end of 2000; 3) our data set is composed by 11 Fama-Bliss yields, Diebold and Li (2006) use 17 Fama-Bliss yields; 4) we have a longer sample. We compute 129 1-month ahead forecasts, from 1993:1 to 2003:9, 124 6-months ahead forecasts from 1993:6 to 2003:9, and 112 and 106 forecasts for 12 and 24 months ahead respectively. Diebold and Li (2006) compute 83 forecasts for the period 1994:2 - 2000:12.

It turns out that the sample is responsible for the differences between the VAR(1) and the AR(1) specification. We first replicate the forecast of the three models (random walk, Nelson-Siegel AR(1) and Nelson-Siegel VAR(1)) considered in Diebold and Li (2006) on their sample, using dynamic forecast on a recursive window as they do. We then extend the forecast period from 2000:12 to 2003:9 where our dataset and forecasts end. Table 5 reports the

FRMSE results of the replications and extensions¹¹. The first column shows selected results from Tables 4 to 6 in Diebold and Li (2006) for the three models¹². The second column reports our replication on their sample; the third column shows results on our sample and column 4 reports results from Table 2A. In each column we denote in bold the best forecast among the three models for each yield-horizon combination. We underline the number in the VAR(1) forecast if it is lower than the AR(1) forecast.

The first column shows that the Nelson-Siegel three factor AR(1) model outperforms random walk most of the cases, especially at the 6 and 12 month horizons, and does better than VAR(1) for 12 month horizon. In column 2, we use our 11 yields dataset and replicate their experiment. The random walk forecast replication matches the Diebold-Li results. The Nelson-Siegel AR(1) model matches well for 1 and 6 month horizons with differences up to a few percentage points, the 12-months forecast produce higher RMSE for both the AR(1) and the VAR(1). The conclusion that the AR(1) setting outperforms the random walk in most cases for 6 and 12 month horizons and the VAR(1) forecasts remains valid. We notice that the VAR(1) delivers lower FRMSE for the 3 and 12 month yields at the 1-step ahead horizon.

Moving to column 3 where we extend the forecast horizon to 2003:9, the picture changes dramatically. The AR(1) setting deteriorates, and the random walk forecast dominates for most cases (exceptions are the 3 month short rate at 1 and 6 month horizons where the VAR(1) does the best). The VAR(1) outperforms the AR(1) for the 12 month yields at 1 and 6 months horizons and for 10 year yield at 1 month horizon. Although for the 12 months ahead forecast the VAR(1) still delivers worse forecast than the AR(1), the difference is not as large as in the previous forecast period.

What drives this shift of performance among the three models? We plot three yields (3 month, 3 year and 10 year yields) for the forecast period 1994:2 - 2003:9 in Figure 3. We can see that up to 2000:12, which is the end of the forecast period in Diebold and Li (2006), the yield curve is quite stable. When we look at the extended period from 2001:1 to 2003:9, the yield curve shows a large downward shift, in which the 3 month yield declines from 4.973% to 0.931%. The slope of the yield curve also changes: the difference between the short to long term yields widens from nearly zero to about 300 basis points.

In a relatively tranquil period, the argument in Diebold and Li (2006) favoring the AR specification applies: "unrestricted VARs tend to produce poor

¹¹In Table 5, we show the FRMSE and not the ratio with respect to the random walk.

¹²1 and 6 months ahead forecasts for Nelson-Siegel VAR(1) model are not reported in Diebold-Li (2006).

forecast due to the large number of included parameters and the resulting potential for in-sample overfitting". The richness of the VAR(1) specification becomes useful in the presence of substantial shifts in the term structure.

In conclusion, the two-step OLS estimation procedure for unrestricted Nelson-Siegel VAR(1) models with fixed λ seems to be robust to a series of robustness checks.

7 Conclusions

We propose a general state-space modeling framework to accommodate a vast number of existing yield curve models. Within this framework, we systematically examine the relative importance of no-arbitrage restrictions versus large information set in forecasting the yield curve. The way we conduct our experiment and comparison helps us to reveal a number of interesting aspects of yield curve forecasting.

Large information set models are useful in forecasting, especially at long horizons for long maturities. No-arbitrage models with constant prices of risk improve the overall forecast performance; more complex specifications with time-varying risk prices or exogenous variables in the state vector are associated with higher prediction errors, especially at longer maturities.

We find evidence of important effects of macroeconomic variables on the future yield curve, in line with the findings of Ang and Piazzesi (2003), Diebold, Rudebusch and Aruoba (2006). We find the predictive power of macro information is high at relatively long forecast horizons (1 to 2 years) and for yields with short-to-medium maturities. This effect remains even in models with only macro factors and no yield factors.

We also find that the yield curve is more useful in forecasting real activity than inflation; we also find that, in addition to the level factor of yield, the slope and curvature factors contribute to the forecast of macro variables.

8 References

1. Ang, A. and M. Piazzesi (2003), "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables." *Journal of Monetary Economics*, 50, 4, 745-787.
2. Ang, A., G. Bekaert and M. Wei (2007), "Do Macro Variables, Asset Markets or Surveys Forecast Inflation Better?" *Journal of Monetary Economics*, 54, 1163-1212.

3. Ang, A., M. Piazzesi and M. Wei (2006), "What Does the Yield Curve Tell us about GDP Growth?" *Journal of Econometrics*, 131, 359-403.
4. Bernanke, B. S. and J. Boivin (2003), "Monetary Policy in a Data-Rich Environment." *Journal of Monetary Economics*, 50, 3, 525-546.
5. Bernanke, B. S., P. Elias and J. Boivin (2005), "Measuring Monetary Policy: A Factor Augmented Vector Autoregressive (FAVAR) Approach." *Quarterly Journal of Economics*, 120, 1, 387-422.
6. Chen, R. and L. Scott (1993), "Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates." *Journal of Fixed Income*, 3, 14-31.
7. Christensen, J.H.E., F.X. Diebold and G.D. Rudebusch (2007), "The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models." *NBER Working Paper* No. 13611.
8. Dai, Q. and K. Singleton (2000), "Specification Analysis of Affine Term Structure Models." *Journal of Finance*, 55, 5, 1943-78.
9. Dai, Q. and K. Singleton (2002), "Expectation Puzzles, Time-varying Risk Premia, and Affine Models of the Term Structure." *Journal of Financial Economics*, 63, 415-41.
10. Diebold, F.X. and C. Li (2006), "Forecasting the Term Structure of Government Bond Yields." *Journal of Econometrics*, 130, 337-364.
11. Diebold, F.X., G.D. Rudebusch, and S. B. Aruoba (2006), "The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach." *Journal of Econometrics*, 131, 309-338.
12. Duffee, G.R. (2002), "Term Premia and Interest Rate Forecasts in Affine Models." *Journal of Finance*, 57, 1, 405-443.
13. Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005), "The Generalized Dynamic Factor Model: One-Sided Estimation and Forecasting." *Journal of the American Statistical Association*, 100, 830-40.
14. Giannone, D., L. Reichlin, and L. Sala (2004), "Monetary Policy in Real-Time." *NBER Macroeconomics Annual*.
15. Kim, D.H. (2007), "Challenges in Macro-Finance modeling." *BIS Working Papers* No. 240.

16. Koop, G. (2003), "*Bayesian Econometrics*." Wiley-Interscience.
17. Le Grand, F. (2007), "Nelson and Siegel, No-Arbitrage and Risk Premium." Paris School of Economics, mimeo.
18. Litterman, R., J. Scheinkman (1991), "Common Factors Affecting Bond Returns." *Journal of Fixed Income*, 1, 54-61.
19. Marcellino, M., J. Stock and M. Watson (2006), "A Comparison of Direct and Iterated AR Methods for Forecasting Macroeconomic Series h-Steps Ahead", CEPR Working Paper No. 4976.
20. Moench, E. (2008), "Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-Augmented VAR Approach." *Journal of Econometrics*, 146, 1, 26-43.
21. Nelson C.R. and A.F. Siegel (1987), "Parsimonious modeling of yield curves." *Journal of Business*, 60, 473-89
22. Ng, S., and S. Ludvigson (2006), "Macro Factors in Bond Risk Premia." Forthcoming in *The Review of Financial Studies*.
23. Pericoli, M. and M. Taboga (2008), "Canonical Term-Structure Models with Observable Factors and the Dynamics of Bond Risk Premia," *Journal of Money, Credit and Banking*, 40, 7, 1471-1488.
24. Piazzesi, M. (2003), "Affine Term Structure Models." prepared for *Handbook of Financial Econometrics*..
25. Rudebusch, G.D. and T. Wu (2008), "A Macro-Finance Model of the Term Structure, Monetary Policy, and the Economy," *Economic Journal*, 118, 906-926.
26. Stock, J.H. and M.W. Watson (1999),. "Forecasting inflation", *Journal of Monetary Economics*, 44, 293–335.
27. Stock, J.H. and M.W. Watson (2002), "Macroeconomic Forecasting Using Diffusion Indexes." *Journal of Business and Economic Statistics*, 20, 147-162.
28. Taylor, J. B. (1993), "Discretion versus Policy Rules in Practice." *Carnegie-Rochester Conference Series on Public Policy*, 39, 195-214.

9 Appendix

9.1 Appendix 1. No-Arbitrage Restrictions on Bond Pricing Parameters

1. State variable dynamics.

Transition equation for X_t follows VAR(1):

$$X_t = \mu + \Phi X_{t-1} + v_t,$$

v_t is i.i.d. $N(0, \Omega)$.

2. Short rate equation.

$$r_t = \delta_0 + \delta_1' X_t$$

δ_0 : a scalar.

δ_1 : $K \times 1$ vector.

3. Time-varying prices of risk (associated with the sources of uncertainty v_t).

$$\Lambda_t = \lambda_0 + \lambda_1 X_t$$

Λ_t : $K \times 1$ vector.

λ_0 : $K \times 1$ vector.

λ_1 : $K \times K$ matrix.

If investors are risk-neutral, $\lambda_0 = 0$ and $\lambda_1 = 0$, hence $\Lambda_t = 0$, no risk adjustment. If $\lambda_0 \neq 0$ and $\lambda_1 = 0$, then price of risk is constant.

4. Pricing kernel.

No arbitrage opportunity between bonds with different maturities implies that there is a discount factor m linking the price of yield of maturity n this month with the yield of maturity $n - 1$ next month.

$$P_t^{(n)} = E_t \left[m_{t+1} P_{t+1}^{(n-1)} \right]$$

The stochastic discount factor is related to the short rate and risk perceived by the market,

$$m_{t+1} = \exp \left(-r_t - \frac{1}{2} \Lambda_t' \Omega \Lambda_t - \Lambda_t' v_{t+1} \right)$$

No-arbitrage recursive relation can be derived from the above equations as:

$$\begin{aligned}
P_t^{(n)} &= E_t \left[m_{t+1} P_{t+1}^{(n-1)} \right] = E_t \left[m_{t+1} m_{t+2} P_{t+2}^{(n-2)} \right] \\
&= E_t \left[m_{t+1} m_{t+2} \dots m_{t+n} P_{t+n}^{(0)} \right] = E_t \left[m_{t+1} m_{t+2} \dots m_{t+n} \cdot 1 \right] \\
&= E_t \left[\exp \left(- \sum_{i=0}^{n-1} \left(r_{t+i} + \frac{1}{2} \Lambda'_{t+i} \Omega \Lambda_{t+i} + \Lambda'_{t+i} v_{t+1+i} \right) \right) \right] \\
&= E_t \left[\exp (A_n + B'_n X_t) \right] = E_t \left[\exp (-n y_{t,n}) \right] \\
&= E_t^Q \left[\exp \left(- \sum_{i=0}^{n-1} r_{t+i} \right) \right]
\end{aligned}$$

E_t^Q denotes the expectation under the risk-neutral probability measure, under which the dynamics of the state vector X_t are characterized by the risk-neutral vector of constants μ^Q and by the autoregressive matrix Φ^Q :

$$\begin{aligned}
\mu^Q &= \mu - \Omega \lambda_0 \\
\Phi^Q &= \Phi - \Omega \lambda_1
\end{aligned}$$

Affine functions of the state variables for yields are:

$$\begin{aligned}
p_{t,t+n} &\equiv \ln P_t^{(n)} = A_n + B'_n X_t \\
y_{t,t+n} &= a_n + b'_n X_t = \frac{-1}{n} (A_n + B'_n X_t)
\end{aligned}$$

where the coefficients follow the difference equations:

$$\begin{aligned}
A_{n+1} &= A_n + B'_n (\mu - \Omega \lambda_0) + \frac{1}{2} B'_n \Omega B_n + A_1 \\
B'_{n+1} &= B'_n (\Phi - \Omega \lambda_1) + B'_1
\end{aligned}$$

with $a_1 = \delta_0 = -A_1$ and $b_1 = \delta_1 = -B_1$ ¹³.

These can be derived from the pricing kernel equation.

$$\begin{aligned}
P_t^{(n+1)} &= E_t \left[m_{t+1} P_{t+1}^{(n)} \right] \\
&= E_t \left[\exp \left\{ -r_t - \frac{1}{2} \Lambda'_t \Omega \Lambda_t - \Lambda'_t v_{t+1} \right\} \exp \{ A_n + B'_n X_{t+1} \} \right] \\
&= \exp \left\{ -r_t - \frac{1}{2} \Lambda'_t \Omega \Lambda_t + A_n \right\} E_t \left[\exp \{ -\Lambda'_t v_{t+1} + B'_n X_{t+1} \} \right] \\
&= \exp \left\{ -\delta_0 - \delta'_1 X_t - \frac{1}{2} \Lambda'_t \Omega \Lambda_t + A_n \right\} \\
&\quad \cdot E_t \left[\exp \{ -\Lambda'_t v_{t+1} + B'_n (\mu + \Phi X_t + v_{t+1}) \} \right]
\end{aligned}$$

¹³Differently from Ang and Piazzesi (2003) where they define $\Sigma \Sigma' = \Omega$ in the difference equations above, and identify the matrix Σ ; we are only interested in Ω . Our λ_0 and λ_1 have therefore different meaning and scale to theirs. The pricing kernel we specify is: $m_{t+1} = \exp(-r_t - \frac{1}{2} \Lambda'_t \Omega \Lambda_t - \Lambda'_t v_{t+1})$, where $v_{t+1} \sim N(0, \Omega)$, while they assume $m_{t+1} = \exp(-r_t - \frac{1}{2} \Lambda'_t \Lambda_t - \Lambda'_t \varepsilon_{t+1})$, where $\varepsilon_{t+1} \sim N(0, I)$

$$\begin{aligned}
&= \exp \left\{ -\delta_0 - \delta'_1 X_t - \frac{1}{2} \Lambda'_t \Omega \Lambda_t + A_n + B'_n (\mu + \Phi X_t) \right\} \\
&\quad \cdot E_t \left[\exp \left\{ -\Lambda'_t v_{t+1} + B'_n v_{t+1} \right\} \right] \\
&= \exp \left\{ -\delta_0 + A_n + B'_n \mu + (B'_n \Phi - \delta'_1) X_t - \frac{1}{2} \Lambda'_t \Omega \Lambda_t \right\} \\
&\quad \cdot E_t \left[\exp \left\{ (-\Lambda'_t + B'_n) v_{t+1} \right\} \right] \\
&= \exp \left\{ -\delta_0 + A_n + B'_n \mu + (B'_n \Phi - \delta'_1) X_t - \frac{1}{2} \Lambda'_t \Omega \Lambda_t \right\} \\
&\quad \cdot \exp \left\{ E_t \left[(-\Lambda'_t + B'_n) v_{t+1} \right] + \frac{1}{2} \text{var} \left[(-\Lambda'_t + B'_n) v_{t+1} \right] \right\} \\
&= \exp \left\{ -\delta_0 + A_n + B'_n \mu + (B'_n \Phi - \delta'_1) X_t - \frac{1}{2} \Lambda'_t \Omega \Lambda_t \right\} \\
&\quad \exp \left\{ \frac{1}{2} \text{var} \left[(-\Lambda'_t + B'_n) v_{t+1} \right] \right\} \\
&= \exp \left\{ -\delta_0 + A_n + B'_n \mu + (B'_n \Phi - \delta'_1) X_t - \frac{1}{2} \Lambda'_t \Omega \Lambda_t \right\} \\
&\quad \cdot \exp \left\{ \frac{1}{2} E_t \left[(-\Lambda'_t + B'_n) v_{t+1} v'_{t+1} (-\Lambda_t + B_n) \right] \right\} \\
&= \exp \left\{ -\delta_0 + A_n + B'_n \mu + (B'_n \Phi - \delta'_1) X_t - \frac{1}{2} \Lambda'_t \Omega \Lambda_t \right\} \\
&\quad \cdot \exp \left\{ \frac{1}{2} \left[\Lambda'_t \Omega \Lambda_t - 2 B'_n \Omega \Lambda_t + B'_n \Omega B_n \right] \right\} \\
&= \exp \left\{ -\delta_0 + A_n + B'_n \mu + (B'_n \Phi - \delta'_1) X_t - B'_n \Omega \Lambda_t + \frac{1}{2} B'_n \Omega B_n \right\} \\
&= \exp \left\{ -\delta_0 + A_n + B'_n \mu + (B'_n \Phi - \delta'_1) X_t - B'_n \Omega (\lambda_0 + \lambda_1 X_t) + \frac{1}{2} B'_n \Omega B_n \right\} \\
&= \exp \left\{ -\delta_0 + A_n + B'_n (\mu - \Omega \lambda_0) + \frac{1}{2} B'_n \Omega B_n + (B'_n \Phi - B'_n \Omega \lambda_1 - \delta'_1) X_t \right\} \\
&= \exp \left\{ [A_1 + A_n + B'_n (\mu - \Omega \lambda_0) + \frac{1}{2} B'_n \Omega B_n] + [B'_n \Phi - B'_n \Omega \lambda_1 + B'_1] X_t \right\}
\end{aligned}$$

5. An alternative presentation for the no-arbitrage coefficients.

In order to understand intuitively how these restrictions are imposed directly on the coefficients in the yield equation, we can write them in the following affine form.

Given that

$$\begin{aligned}
p_{t,t+n} &= A_n + B'_n X_t \\
y_{t,t+n} &= a_n + b'_n X_t = \frac{-1}{n} (A_n + B'_n X_t)
\end{aligned}$$

we can derive

$$\begin{aligned}
b_{n+1} &= \frac{1}{(n+1)} \left[\sum_{i=0}^n (\Phi' - \lambda'_1 \Omega)^i \right] b_1 \\
a_{n+1} &= a_1 - \frac{1}{(n+1)} \sum_{i=1}^n B^{(i)}
\end{aligned}$$

where $B^{(i)} = B'_i (\mu - \Omega \lambda_0) + \frac{1}{2} B'_i \Omega B_i$.

9.2 Appendix 2. The likelihood function with Chen-Scott (1993) method

In order to be able to extract factors under no-arbitrage restrictions, we employ the method by Chen and Scott (1993)¹⁴. We assume that there are K factors in the state equation and that among them, K_2 factors are unobserved. When the number of yields N exceeds number of unobserved factors K_2 , we assume that K_2 yields, y_t^{NE} , are observed without measurement errors, and that $N - K_2$ yields, y_t^E , are measured with error u_t^m . The state vector contains both observed variables X_t^o and latent factors X_t^u , thus $X_t = [X_t^o; X_t^u]$

The measurement equation can be written as follows:

$$y_t = a + b^o X_t^o + b^u X_t^u + b^m u_t^m$$

$$\text{where } y_t = \begin{bmatrix} y_t^{NE} \\ y_t^E \end{bmatrix}, a = \begin{bmatrix} a^{NE} \\ a^E \end{bmatrix}, b^o = \begin{bmatrix} b^{NE,o} \\ b^{E,o} \end{bmatrix}, b^u = \begin{bmatrix} b^{NE,u} \\ b^{E,u} \end{bmatrix}, \text{ and } b^m = \begin{bmatrix} \mathbf{0}_{(K_2 \times (N-K_2))} \\ b^{E,m} \end{bmatrix}$$

For a given parameter vector $\theta = (\mu, \Phi, \Omega, \delta_0, \delta_1, \lambda_0, \lambda_1)$, the unobserved factors X_t^u will be solved from the yields and the observed variables X_t^o as: $X_t^u = (b^{NE,u})^{-1} [Y_t^{NE} - a^{NE} - b^{NE,o} X_t^o]$.

Denoting the normal density functions of the state variables X_t^u and the error u_t^m as f_X and f_{u^m} respectively, the joint likelihood $\mathcal{L}(\theta)$ of the observed data on zero coupon yields Y_t and the observable factors X_t^o is given (up to a constant) by:

$$\begin{aligned} \mathcal{L}(\theta) &= \prod_{t=2}^T f(y_t, X_t^o | y_{t-1}, X_{t-1}^o) \\ \log(\mathcal{L}(\theta)) &= \sum_{t=2}^T \log |\det(J^{-1})| + \log f_X(X_t^o, X_t^u | X_{t-1}^o, X_{t-1}^u) + \log f_{u^m}(u_t^m) \\ &= -(T-1) \log |\det(J)| - \frac{(T-1)}{2} \log(\det(\Omega)) \\ &\quad - \frac{1}{2} \sum_{t=2}^T (X_t - \mu - \Phi X_{t-1})' \Omega^{-1} (X_t - \mu - \Phi X_{t-1}) \\ &\quad - \frac{(T-1)}{2} \sum_{i=1}^{N-K_2} \log(\sigma_i^2) - \frac{1}{2} \sum_{t=2}^T \sum_{i=1}^{N-K_2} \frac{(u_{t,i}^m)^2}{\sigma_i^2} \end{aligned}$$

$$\text{The Jacobian term is: } J = \begin{pmatrix} I_{K-K_2} & 0_{(K-K_2) \times K_2} & 0_{(K-K_2) \times (N-K_2)} \\ B^o & B^u & B^m \end{pmatrix}.$$

¹⁴Our presentation follows closely the discussion in Ang and Piazzesi (2003).

Figure 1: Macro variables

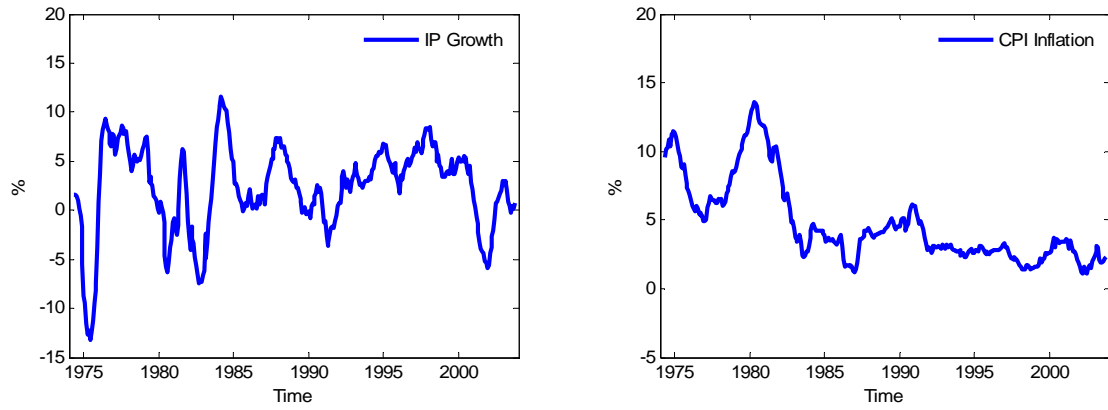


Figure 2: First four macro factor

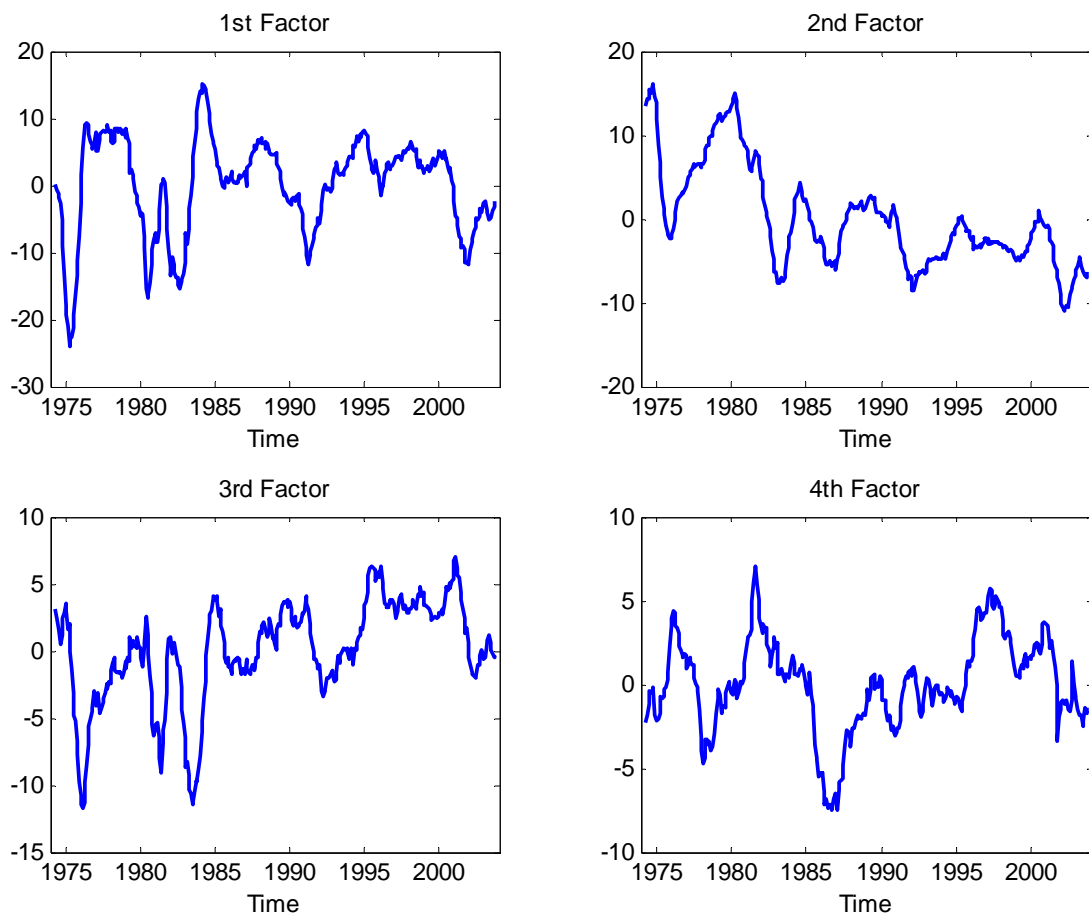
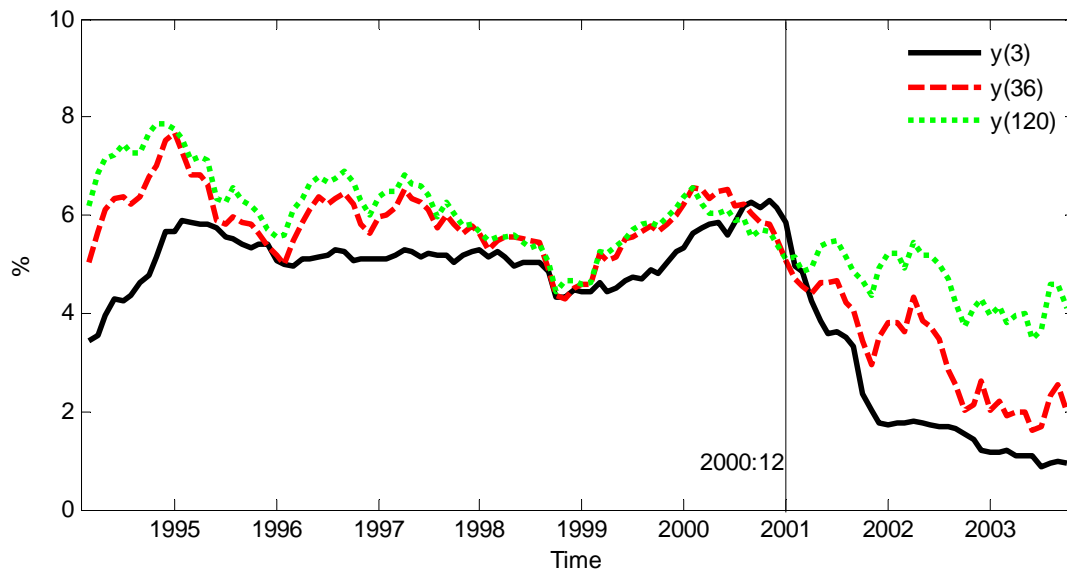


Figure 3: Forecast periods comparison



Notes:

- 1) This figure shows three yields (3 month, 3 year and 10 year) in the period from 1994:2 to 2003:9.
- 2) In the Diebold-Li (2006) paper, forecast for yields are made in the period between 1994:2- 2000:12. The straight line denotes the time of 2000:12 when their forecast ends.
- 3) Starting from 2001:1 to 2003:9, the yield curve has experienced a large downward movement where the 3-month yield declined from 4.973% in 2001:1 to 0.931% in 2003:9, a total change of more than 400 basis points. The medium and long term interests rates also fall dramatically.

Table 1 Factor loadings

Factor 1	Total variance explained: 32.26%	R^2
	Index of IP: Non-energy, total	0.93
	Index of IP: Mfg	0.93
	Index of IP: Non-energy excl CCS	0.92
	Index of IP: Total	0.92
	Index of IP: Non-energy excl CCS and MVP	0.91
Factor 2	Total variance explained: 22.25%	
	PPI: crude materials	0.80
	CPI: housing	0.79
	CPI: services	0.79
	Loans and Securities @ all commercial banks: commercial and Individual loans (in mil of current \$)	0.78
	CPI: food and beverages	0.77
Factor 3	Total variance explained: 9.04%	
	Loans and Securities @ all commercial banks: Securities, U.S. govt (in mil of current \$)	0.55
	Loans and Securities @ all commercial banks: Securities, total (in mil of current \$)	0.41
	ISM mfg index: Employment	0.39
	M1 (in bil of current \$)	0.38
	Mean duration of unemployment	0.38
Factor 4	Total variance explained: 4.69%	
	Nominal effective exchange rate	0.65
	Spot Euro/US	0.63
	Spot SZ/US	0.56
	Depository institutions reserve: Total (adj for rr changes)	0.35
	Spot Japan/US	0.34
Factor 5	Total variance explained: 3.97%	
	M3 (in bil of current \$)	0.40
	M2 (in bil of current \$)	0.31
	Employment on nonag payrolls: Financial activities	0.21
	Loans and Securities @ all commercial banks: Total (in mil of current \$)	0.19
	Total merchandise exports (FAS value) (in mil of \$)	0.19

Notes: Factors are extracted from a panel with 162 macro variables (1974:4-B003:9) after transformation to control for stationarity. The first five factors are shown with the five variables with which they are most highly correlated. The first five factors together explain 72.21% of the total variation in the transformed macro panel. The series used are the same as in Giannone, Reichlin and Sala (2004), except that we exclude 9 interest rate variables.

Notes on the forecast comparison for Table 2 to Table 4:

- 1) Rolling forecast. We set the estimation window fixed at 180 periods. By moving the sample forward one observation at a time, we implement rolling estimation starting from the sample period of 1978:1-1992:12.
- 2) $y(n)$: yield of maturity n , n measured in month;
 h : forecast horizon in month;
 X : $\{X_1 X_2 X_3\}$, the state vector and the variables.
 NS : Nelson-Siegel yield factors.
 CS : Chen-Scott yield factors.
 mf : macro factors.
- 3) Forecasting period. For 1 month ahead forecasting horizon, we conduct our exercise for all dates in the period 1993:1 - 2003:9, a total of 129 periods. For 6 month, 18 month and 24 month ahead forecast, we have a total of 124, 112 and 106 forecast periods respectively.
- 4) FRMSE ratio: the first forecast performance measure constructed from forecast root mean squared error (FRMSE) of competing models. For yield forecast in Table 2-A and Table 4, we calculate the ratio of forecast root mean squared error (FRMSE) of the specific model with respect to that of Random Walk for $y(n)$ on forecast horizon h months ahead; for macro variable forecast in Table 3-A, we take the ratio of the FRMSE of the selected model with respect to that of a simple AR(1) model.
- 5) BMAI ratio: the second forecast performance measure constructed from a Bayesian model averaging index (BMAI). For yield forecast in Table 2-B, we calculate the ratio of posterior probability that model i 's forecast $\hat{y}(n)_i$ is included in a realised yield $y(n)$'s regression with respect to that of Random Walk forecast $\hat{y}(n)_{rw}$: $y(n) = \beta_0 + \beta_1 \hat{y}(n)_i + \beta_2 \hat{y}(n)_{rw}$; for macro variable forecast in Table 3-B, we calculate the ratio of the posterior probability of selected model forecast with respect to that of a simple AR(1) model.
- 6) Risk price specifications in no-arbitrage models, Table 3-A and Table 3-B:
 $trp = 0$: constant risk price;
 $trp = 1$: time varying risk price, λ_1 diagonal;
 $trp = 2$: time-varying risk price, λ_1 full matrix.

- 7) Illustration on the display of FRMSE ratio:

< 0.80	0.80 – 0.90	0.90 – 1.00	> 1.00
--------	-------------	-------------	--------

- 8) Illustration on the display of BMAI ratio:

< 0.5	0.5 - 2	2.00 – 5.00	> 5.00
-------	---------	-------------	--------

Table 2. Forecast comparison of yields

Table 2-A. FRMSE ratios

Info.	n. yield	n. macro	Unrestricted	No-arbitrage Restricted	
				Const. risk price	Time-varying risk price, λ_1 diagonal
Small N	3	0	X: {NS1 NS2 NS3} (Diebold-Li)	X: {CS1 CS2 CS3} A0(3)	X: {CS1 CS2 CS3} A0(3)
			y(n)\h	y(n)\h	y(n)\h
			1 6 12 24	1 6 12 24	1 6 12 24
			y(3) 0.89 0.84 0.88 0.93	y(3) 0.91 0.85 0.91 0.87	y(3) 0.84 0.75 0.84 0.90
			y(12) 1.06 0.98 0.95 0.96	y(12) 1.03 0.97 0.96 0.84	y(12) 0.99 0.95 0.94 0.96
			y(36) 1.02 1.01 0.98 1.06	y(36) 0.98 0.98 0.95 0.84	y(36) 1.01 1.02 1.00 1.12
			y(60) 1.04 1.03 1.00 1.17	y(60) 1.02 0.99 0.96 0.87	y(60) 1.06 1.06 1.05 1.30
			y(120) 1.06 1.06 1.06 1.37	y(120) 1.00 1.00 0.98 0.93	y(120) 1.02 1.09 1.14 1.62
		2	X: {NS1 NS2 NS3 infl IPgrowth}	X: {CS1 CS2 CS3 infl IPgrowth}	X: {CS1 CS2 CS3 infl IPgrowth}
			y(n)\h	y(n)\h	y(n)\h
			1 6 12 24	1 6 12 24	1 6 12 24
			y(3) 0.89 0.87 0.92 0.93	y(3) 0.89 0.81 0.88 0.84	y(3) 0.85 0.78 0.86 0.87
			y(12) 1.03 1.00 0.97 0.96	y(12) 0.96 0.95 0.95 0.82	y(12) 0.95 0.97 0.96 0.91
Large N	3	2	X: {NS1 NS2 NS3 mf1 mf2}	X: {CS1 CS2 CS3 mf1 mf2}	X: {CS1 CS2 CS3 mf1 mf2}
			y(n)\h	y(n)\h	y(n)\h
			1 6 12 24	1 6 12 24	1 6 12 24
			y(3) 0.89 0.85 0.87 0.85	y(3) 0.92 0.87 0.93 0.88	y(3) 0.84 0.77 0.79 0.83
			y(12) 1.01 0.96 0.91 0.88	y(12) 1.00 0.96 0.95 0.83	y(12) 0.98 0.93 0.85 0.87
			y(36) 1.05 1.07 1.01 1.05	y(36) 1.00 0.97 0.95 0.80	y(36) 1.04 1.05 0.97 0.98
			y(60) 1.07 1.12 1.08 1.24	y(60) 1.03 0.98 0.94 0.80	y(60) 1.09 1.14 1.10 1.16
			y(120) 1.08 1.17 1.19 1.56	y(120) 1.00 0.98 0.96 0.88	y(120) 1.06 1.27 1.34 1.58
		4	X: {NS1 NS2 NS3 mf1 mf2 mf3 mf4}	X: {y(1) mf1 mf2 mf3 mf4} (Moench)	X: {y(1) mf1 mf2 mf3 mf4} (Moench)
			y(n)\h	y(n)\h	y(n)\h
			1 6 12 24	1 6 12 24	1 6 12 24
			y(3) 0.93 0.84 0.77 0.69	y(3) 1.92 1.09 0.87 0.72	y(3) 1.74 1.06 0.88 0.73
			y(12) 1.02 0.89 0.79 0.73	y(12) 3.32 1.25 0.88 0.79	y(12) 2.28 1.11 0.87 0.76
	1	4	y(36) 1.03 0.97 0.89 0.84	y(36) 4.09 1.65 1.26 1.26	y(36) 2.63 1.25 0.94 0.87
			y(60) 1.05 1.03 0.96 0.97	y(60) 5.22 2.10 1.56 1.63	y(60) 3.19 1.41 1.03 1.07
			y(120) 1.05 1.08 1.08 1.18	y(120) 6.69 2.80 2.10 2.50	y(120) 4.10 1.83 1.38 1.63
			X: {y(1) mf1 mf2 mf3 mf4}	X: {mf1 mf2 mf3 mf4}	X: {mf1 mf2 mf3 mf4}
			y(n)\h	y(n)\h	y(n)\h
			1 6 12 24	1 6 12 24	1 6 12 24
	0	4	y(3) 1.53 1.03 0.86 0.72	y(3) 5.64 1.47 0.83 0.72	y(3) 5.32 1.40 0.81 0.76
			y(12) 2.09 1.05 0.83 0.74	y(12) 4.82 1.41 0.83 0.84	y(12) 4.69 1.35 0.80 0.86
			y(36) 2.37 1.12 0.83 0.84	y(36) 4.04 1.58 1.23 1.28	y(36) 4.22 1.48 0.95 1.12
			y(60) 2.53 1.17 0.88 0.98	y(60) 4.93 1.98 1.49 1.68	y(60) 4.30 1.62 1.09 1.40
			y(120) 2.82 1.36 1.09 1.35	y(120) 6.08 2.59 1.97 2.45	y(120) 4.90 2.07 1.51 2.03
			X: {mf1 mf2 mf3 mf4}	X: {mf1 mf2 mf3 mf4}	X: {mf1 mf2 mf3 mf4}
			y(n)\h	y(n)\h	y(n)\h
			1 6 12 24	1 6 12 24	1 6 12 24
			y(3) 5.38 1.42 0.81 0.71	y(3) 5.64 1.47 0.83 0.72	y(3) 5.32 1.40 0.81 0.76
			y(12) 4.53 1.32 0.77 0.76	y(12) 4.82 1.41 0.83 0.84	y(12) 4.69 1.35 0.80 0.86
			y(36) 3.84 1.29 0.79 0.89	y(36) 4.04 1.58 1.23 1.28	y(36) 4.22 1.48 0.95 1.12
			y(60) 3.70 1.30 0.84 1.05	y(60) 4.93 1.98 1.49 1.68	y(60) 4.30 1.62 1.09 1.40
			y(120) 3.64 1.47 1.08 1.42	y(120) 6.08 2.59 1.97 2.45	y(120) 4.90 2.07 1.51 2.03

Table 2-B. BMAI ratios

Info.	n. yield	n. macro	Unrestricted	No-arbitrage Restricted													
				Const. risk price								Time-varying risk price, λ_1 diagonal					
Small N	3	0	X: {NS1 NS2 NS3} (Diebold-Li)					X: {CS1 CS2 CS3} A0(3)					X: {CS1 CS2 CS3} A0(3)				
			y(n)\h	1	6	12	24	y(n)\h	1	6	12	24	y(n)\h	1	6	12	24
			y(3)	0.9	1.0	1.0	1.0	y(3)	1.2	0.8	1.1	1.0	y(3)	0.9	1.0	1.1	1.5
			y(12)	1.0	1.0	1.0	1.0	y(12)	1.0	1.0	1.0	1.0	y(12)	0.9	1.0	1.0	1.4
			y(36)	1.1	1.0	1.0	1.0	y(36)	1.0	1.0	1.0	1.0	y(36)	1.0	1.0	1.1	3.1
			y(60)	1.0	1.0	1.0	2.4	y(60)	1.0	1.0	1.1	1.1	y(60)	1.0	1.0	2.5	9.5
			y(120)	1.9	1.8	1.2	3.0	y(120)	1.2	1.9	2.1	3.5	y(120)	1.5	2.5	3.2	0.4
		2	X: {NS1 NS2 NS3 infl IPgrowth}					X: {CS1 CS2 CS3 infl IPgrowth}					X: {CS1 CS2 CS3 infl IPgrowth}				
			y(n)\h	1	6	12	24	y(n)\h	1	6	12	24	y(n)\h	1	6	12	24
			y(3)	1.0	1.0	1.0	1.5	y(3)	1.0	0.8	1.1	1.0	y(3)	0.9	0.9	1.8	5.8
			y(12)	1.0	1.0	1.0	1.9	y(12)	0.9	1.0	1.0	1.0	y(12)	1.0	1.0	2.6	2.7
			y(36)	1.0	1.0	1.0	7.8	y(36)	1.0	1.0	1.0	1.0	y(36)	1.0	1.0	2.9	9.0
			y(60)	1.0	1.0	1.0	4.8	y(60)	1.0	1.0	1.0	1.0	y(60)	1.0	1.0	7.0	4.8
			y(120)	1.8	1.0	6.0	0.3	y(120)	1.2	1.9	2.7	3.4	y(120)	1.0	0.6	0.3	0.2
Large N	3	2	X: {NS1 NS2 NS3 mf1 mf2}					X: {CS1 CS2 CS3 mf1 mf2}					X: {CS1 CS2 CS3 mf1 mf2}				
			y(n)\h	1	6	12	24	y(n)\h	1	6	12	24	y(n)\h	1	6	12	24
			y(3)	0.9	0.9	1.3	1.1	y(3)	1.1	1.0	1.4	1.0	y(3)	1.0	0.8	2.0	9.8
			y(12)	1.0	1.0	1.4	1.0	y(12)	1.0	1.0	1.1	1.0	y(12)	0.9	1.1	4.6	12.2
			y(36)	1.0	1.0	2.0	3.7	y(36)	1.0	1.0	1.6	1.4	y(36)	1.0	1.0	10.1	1.6
			y(60)	1.0	1.0	4.7	7.1	y(60)	1.0	1.0	3.9	2.1	y(60)	1.0	1.2	3.5	0.2
			y(120)	1.6	3.5	2.8	0.7	y(120)	1.2	2.0	1.2	4.9	y(120)	0.9	0.3	0.1	0.4
		4	X: {NS1 NS2 NS3 mf1 mf2 mf3 mf4}														
			y(n)\h	1	6	12	24										
			y(3)	1.0	1.0	5.8	16.6										
			y(12)	1.0	1.0	13.0	16.7										
			y(36)	1.0	1.0	8.0	5.1										
			y(60)	1.0	1.0	1.2	1.6										
			y(120)	1.7	4.6	0.2	1.0										
	1	4	X: {y(1) mf1 mf2 mf3 mf4}					X: {y(1) mf1 mf2 mf3 mf4} (Moench)					X: {y(1) mf1 mf2 mf3 mf4} (Moench)				
			y(n)\h	1	6	12	24	y(n)\h	1	6	12	24	y(n)\h	1	6	12	24
			y(3)	1.0	1.0	1.2	11.6	y(3)	1.0	1.0	1.5	10.1	y(3)	1.0	1.0	1.2	10.4
			y(12)	1.0	1.0	1.5	13.3	y(12)	1.0	1.0	7.8	12.3	y(12)	1.0	1.0	2.3	15.5
			y(36)	2.0	1.0	7.1	12.0	y(36)	1.0	3.8	11.5	14.4	y(36)	1.9	1.0	11.9	11.9
			y(60)	3.1	1.7	5.6	10.9	y(60)	1.8	9.7	1.2	5.5	y(60)	0.8	3.5	2.5	4.1
			y(120)	0.2	4.1	0.9	1.6	y(120)	9.7	1.1	1.0	1.0	y(120)	0.3	9.5	0.9	1.0
	0	4	X: {mf1 mf2 mf3 mf4}					X: {mf1 mf2 mf3 mf4}					X: {mf1 mf2 mf3 mf4}				
			y(n)\h	1	6	12	24	y(n)\h	1	6	12	24	y(n)\h	1	6	12	24
			y(3)	1.0	1.0	1.3	9.1	y(3)	1.0	1.0	1.0	14.0	y(3)	1.0	1.0	1.1	6.4
			y(12)	1.0	1.0	1.7	10.8	y(12)	1.0	1.0	2.7	15.1	y(12)	1.0	1.0	1.4	11.4
			y(36)	1.0	1.1	12.3	13.0	y(36)	1.1	5.8	12.1	13.8	y(36)	1.0	2.4	11.7	13.9
			y(60)	1.3	5.7	2.0	12.1	y(60)	3.1	9.7	2.0	13.2	y(60)	2.0	9.7	1.1	5.9
			y(120)	7.3	4.2	0.7	1.1	y(120)	1.9	0.9	1.0	1.1	y(120)	1.6	0.8	0.7	1.0

Table 3. Forecast comparison of macro variables

Table 3-A. FRMSE ratios

Macro	Unrestricted					Restricted				
	3 Nelson-Siegel factors					1 yield factor: 1-month rate				
Inflation	X: {NS1 NS2 NS3 infl.}					X: {CS1 CS2 CS3 infl.}				
	h	1	6	12	24	h	1	6	12	24
		1.02	1.04	1.06	1.24	trp=0	1.02	1.01	0.98	1.09
						trp=1	1.02	1.01	0.95	0.98
						trp=2	1.02	1.09	1.16	1.47
IP growth	X: {NS1 NS2 NS3 infl. IPg}					X: {m1 infl. IPg}				
	h	1	6	12	24	h	1	6	12	24
		1.07	1.28	1.41	1.70		1.03	1.14	1.33	1.29

Table 3-B. BMAI ratios

Macro	Unrestricted					Restricted				
	3 Nelson-Siegel factors					1 yield factor: 1-month rate				
Inflation	X: {NS1 NS2 NS3 infl.}					X: {CS1 CS2 CS3 infl.}				
	h	1	6	12	24	h	1	6	12	24
		0.8	0.9	0.8	0.9	trp=0	0.8	1.0	0.8	5.0
						trp=1	0.8	1.0	0.5	1.1
						trp=2	0.8	0.9	0.4	4.2
IP growth	X: {NS1 NS2 NS3 infl. IPg}					X: {m1 infl. IPg}				
	h	1	6	12	24	h	1	6	12	24
		0.8	0.9	0.8	0.9		1.0	0.5	0.5	0.2

**Table 4. FRMSE ratios for Nelson-Siegel factor models:
1-step MLE estimation with Kalman filter**

Table 4-A. Models with 3 Nelson-Siegel dynamic factors

X: {NS1 NS2 NS3}

Model Setting	Diebold-Li					No-arbitrage Nelson-Siegel				
λ fixed	$y(n) \backslash h$	1	6	12	24	$y(n) \backslash h$	1	6	12	24
	y (3)	0.90	0.86	0.89	0.95	y (3)	0.91	0.88	0.90	0.95
	y (12)	1.07	1.00	0.95	0.97	y (12)	1.09	1.02	0.95	0.96
	y (36)	1.02	1.02	0.97	1.04	y (36)	1.03	1.03	0.96	1.01
	y (60)	1.05	1.03	0.97	1.11	y (60)	1.07	1.04	0.96	1.07
	y (120)	1.07	1.04	0.99	1.24	y (120)	1.10	1.05	0.96	1.14
λ free	$y(n) \backslash h$	1	6	12	24	$y(n) \backslash h$	1	6	12	24
	y (3)	0.89	0.84	0.88	0.93	y (3)	0.91	0.86	0.88	0.93
	y (12)	1.03	0.98	0.95	0.96	y (12)	1.06	1.00	0.95	0.95
	y (36)	1.05	1.01	0.97	1.05	y (36)	1.06	1.02	0.96	1.01
	y (60)	1.06	1.03	0.99	1.13	y (60)	1.09	1.04	0.97	1.08
	y (120)	1.10	1.06	1.04	1.28	y (120)	1.13	1.06	1.00	1.15

Table 4-B. Models with 3 Nelson-Siegel factors and macro variables/factors

λ fixed at 0.0609

X: {NS1 NS2 NS3 infl IPgrowth}					X: {NS1 NS2 NS3 mf1 mf2}					X: {NS1 NS2 NS3 mf1 mf2 mf3 mf4}				
$y(n) \backslash h$	1	6	12	24	$y(n) \backslash h$	1	6	12	24	$y(n) \backslash h$	1	6	12	24
y (3)	0.90	0.87	0.90	0.89	y (3)	0.89	0.88	0.92	0.88	y (3)	0.99	0.85	0.80	0.78
y (12)	1.04	1.01	0.96	0.91	y (12)	1.01	0.99	0.94	0.88	y (12)	1.01	0.92	0.88	0.81
y (36)	1.06	1.11	1.05	1.08	y (36)	1.07	1.10	1.02	1.05	y (36)	1.02	1.01	1.01	0.88
y (60)	1.09	1.16	1.12	1.26	y (60)	1.09	1.15	1.10	1.25	y (60)	1.06	1.07	1.09	0.94
y (120)	1.11	1.21	1.23	1.58	y (120)	1.11	1.20	1.23	1.60	y (120)	1.07	1.15	1.22	1.07

Notes:

- 1) The above table reports 1-step MLE estimation of models with 3 Nelson-Siegel dynamic factors. In Table 4-A, the first column is unrestricted Diebold-Li model, while the second column is No-arbitrage Nelson-Siegel model. The first row is the result when λ fixed at 0.0609, and the second row reports the result when we estimate λ as free parameter.
- 2) In Table 4-A, Model in the first row and first column has the same specification as the Diebold-Li model reported in Table 2. The only difference is the estimation procedure. In Table 2, we estimate the model with two step OLS. We first estimate the Nelson-Siegel factors from yields, then we estimate the VAR(1) dynamics of Nelson-Siegel factors. Here, we explore the state-space framework to estimate the model in one step with Maximum Likelihood, extracting the Nelson-Siegel factors by Kalman Filter.
- 3) In Table 4-B, these three models corresponds to the second to the fourth models in the first column of Table 2, i.e. three Nelson-Siegel factors with macro variables/factors. In Table 2, we estimate these models with two step OLS. We first estimate the Nelson-Siegel factors from yields, then we estimate the VAR(1) dynamics of Nelson-Siegel factors together with macro variable/factors. Here, we explore the state-space framework to estimate the model in one step with Maximum Likelihood, extracting the Nelson-Siegel factors by Kalman Filter and restrict the factor loadings on the measurement equations of yields to be zero.

Table 5. Forecast periods and FRMSE performance between models

Models	Diebold-Li (2006) paper results				DL Forecast replication				DL Forecast extension				Forecast in our paper			
	Forecast period: 1994:2 to 2000:12				Forecast period: 1994:2 to 2000:12				Forecast period: 1994:2 to 2003:9				Forecast period: 1993:1+h-1 to 2003:9			
Random Walk	y (n) \h	1	6	12	y (n) \h	1	6	12	y (n) \h	1	6	12	y (n) \h	1	6	12
	y (3)	0.179	0.605	1.019	y (3)	0.178	0.606	1.019	y (3)	0.214	0.836	1.429	y (3)	0.205	0.809	1.417
	y (12)	0.241	0.779	1.197	y (12)	0.240	0.779	1.195	y (12)	0.259	0.906	1.493	y (12)	0.251	0.877	1.480
	y (36)	0.279	0.879	1.237	y (36)	0.278	0.879	1.237	y (36)	0.309	0.917	1.354	y (36)	0.302	0.891	1.345
	y (60)	0.276	0.861	1.191	y (60)	0.276	0.861	1.190	y (60)	0.307	0.869	1.230	y (60)	0.302	0.850	1.223
	y (120)	0.254	0.758	1.052	y (120)	0.254	0.758	1.048	y (120)	0.282	0.727	1.002	y (120)	0.278	0.719	1.000
3 NS AR(1)	y (n) \h	1	6	12	y (n) \h	1	6	12	y (n) \h	1	6	12	y (n) \h	1	6	12
	y (3)	0.176	0.517	0.739	y (3)	0.183	0.547	0.801	y (3)	0.225	0.927	1.490	y (3)	0.216	0.842	1.395
	y (12)	0.236	0.669	0.841	y (12)	0.243	0.661	0.832	y (12)	0.294	1.023	1.565	y (12)	0.287	0.960	1.466
	y (36)	0.279	0.750	0.918	y (36)	0.279	0.767	1.018	y (36)	0.318	0.980	1.472	y (36)	0.319	0.943	1.326
	y (60)	0.292	0.777	0.978	y (60)	0.286	0.808	1.137	y (60)	0.323	0.936	1.392	y (60)	0.325	0.903	1.223
	y (120)	0.260	0.721	0.981	y (120)	0.251	0.748	1.175	y (120)	0.293	0.789	1.234	y (120)	0.290	0.757	1.033
3 NS VAR(1)	y (n) \h	1	6	12	y (n) \h	1	6	12	y (n) \h	1	6	12	y (n) \h	1	6	12
	y (3)	n. a.	n. a.	1.102	y (3)	<u>0.165</u>	0.582	1.147	y (3)	<u>0.186</u>	<u>0.779</u>	1.591	y (3)	<u>0.182</u>	<u>0.678</u>	<u>1.248</u>
	y (12)	n. a.	n. a.	1.293	y (12)	<u>0.229</u>	0.829	1.387	y (12)	<u>0.262</u>	<u>1.011</u>	1.846	y (12)	<u>0.265</u>	<u>0.864</u>	<u>1.399</u>
	y (36)	n. a.	n. a.	1.393	y (36)	0.299	0.990	1.549	y (36)	0.322	1.081	1.827	y (36)	<u>0.309</u>	<u>0.904</u>	<u>1.315</u>
	y (60)	n. a.	n. a.	1.385	y (60)	0.301	1.009	1.562	y (60)	0.325	1.050	1.700	y (60)	<u>0.313</u>	<u>0.879</u>	1.224
	y (120)	n. a.	n. a.	1.279	y (120)	0.254	0.901	1.461	y (120)	<u>0.292</u>	0.888	1.449	y (120)	0.294	0.761	1.060

Notes:

- 1) In this table, we report the value of FRMSE instead of the ratio between two competing models, in order to compare with the results of Diebold-Li (2006) to understand the different performance of AR(1) vs. VAR(1) settings of Nelson-Siegel dynamic factor models.
- 2) In each column, the three models are forecast with the same data, same forecast strategy and period. The first column compiles the results reported in Diebold-Li (2006) where they make the forecast with a data set of 17 Fama-Bliss yields.
- 3) Among the three models in each column, the lowest FRMSE for a yield-horizon combination is marked with bold face.
- 4) For Nelson-Siegel three factor models, the result is underlined if the VAR(1) forecast better than AR(1) model for a yield-horizon combination.
- 5) Forecast in the first three columns are made with recursive window starting from 1985:1 and dynamic forecast, while the last column presents our results with rolling window of 180 months and iterated forecast.